

Original citation:

Wilhelm, Thomas and Jones, Alexandra M.. (2014) Identification of related peptides through the analysis of fragment ion mass shifts. Journal of Proteome Research . ISSN 1535-3893

Permanent WRAP url:

<http://wrap.warwick.ac.uk/62125>

Copyright and reuse:

The Warwick Research Archive Portal (WRAP) makes this work by researchers of the University of Warwick available open access under the following conditions. Copyright © and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable the material made available in WRAP has been checked for eligibility before being made available.

Copies of full items can be used for personal research or study, educational, or not-for profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

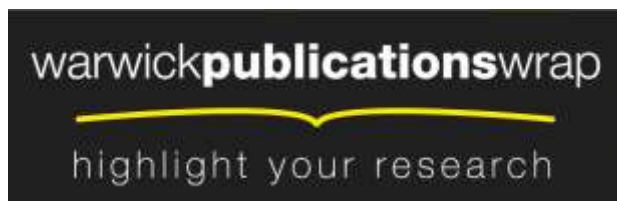
Publisher's statement:

"This document is the Accepted Manuscript version of a Published Work that appeared in final form in Journal of Proteome Research, copyright © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <http://dx.doi.org/10.1021/pr500347e>."

A note on versions:

The version presented here may differ from the published version or, version of record, if you wish to cite this item you are advised to consult the publisher's version. Please see the 'permanent WRAP url' above for details on accessing the published version and note that access may require a subscription.

For more information, please contact the WRAP Team at: publications@warwick.ac.uk



<http://wrap.warwick.ac.uk>

Identification of related peptides through the analysis of fragment ion mass shifts

Thomas Wilhelm^{†}, Alexandra M. E. Jones^{§†}*

* to whom correspondence should be addressed (Thomas.Wilhelm@ifr.ac.uk Tel.+44 1603 255313 Fax ~507723)

[†]Institute of Food Research, Norwich NR4 7UA, UK

[§]The Sainsbury Laboratory, Norwich Research Park, Norwich NR4 7UH, UK

[†]School of Life Sciences, University of Warwick, Gibbet Hill Road, Coventry CV4 7AL, UK

KEYWORDS

Mass spectrometry, MS/MS, data analysis, proteomics, protein modification

ABSTRACT

Mass spectrometry (MS) has become the method of choice to identify and quantify proteins, typically by fragmenting peptides and inferring protein identification by reference to sequence databases. Well-established programs have largely solved the problem of identifying peptides in complex mixtures. However, to prevent the search space from becoming prohibitively large most search engines need a list of expected modifications. Therefore, unexpected modifications limit both the identification of proteins and peptide-based quantification. We developed Mass Spectrometry-Peak Shift Analysis (MS-PSA) to rapidly identify related spectra in large datasets without reference to databases or specified modifications. Peptide identifications from established tools, such as MASCOT or SEQUEST, may be propagated onto MS-PSA results.

1
2
3 Modification of a peptide alters the mass of the precursor ion and some of the fragmentation
4 ions. MS-PSA identifies characteristic fragmentation masses from MS/MS spectra. Related
5 spectra are identified by pattern matching of unchanged and mass-shifted fragment ions. We
6 illustrate the use of MS-PSA with simple and complex mixtures with both high and low mass
7 accuracy datasets. MS-PSA is not limited to the analysis of peptides but can be used for the
8 identification of related groups of spectra in any set of fragmentation patterns.
9
10
11
12
13
14
15
16
17
18

19 Introduction

20
21 Currently, the most widely used method to identify proteins and their post-translational
22 modifications is through the analysis of enzymatically-derived peptides by tandem mass
23 spectrometry. Mass spectrometers coupled to liquid chromatography systems are capable of
24 routinely analysing complex mixtures of peptides and may generate tens of thousands of mass
25 spectra per hour. Typically only a small fraction of the spectra acquired contribute to the
26 identification of peptides. Part of this inefficiency is due to acquisition of spectra when peptides
27 are not eluting, limitations of sensitivity and sample contamination with non-peptide molecules.
28 However, many unidentified spectra remain that show characteristics of peptide fragmentation,
29 such as regularly spaced ions.
30
31
32
33
34
35
36
37
38
39
40
41

42 Computational methods for assigning peptide identifications to MS/MS spectra can be grouped
43 into two main categories: database similarity searches and de novo techniques, which directly
44 reconstruct peptide sequences from spectra. Database searches can be further subdivided into
45 two different approaches that use either theoretical spectra predicted from peptide sequences or
46 empirical high quality spectra (spectral libraries). Well-known search engines that use theoretical
47 spectra are SEQUEST¹, MASCOT², X!Tandem³, and OMSSA⁴, amongst others. However, due
48 to insufficient understanding of the factors that determine peptide fragmentation, most current
49
50
51
52
53
54
55
56
57
58
59
60

search tools employ simplified fragmentation models, such as the uniform backbone dissociation model, leading to many unidentified or misidentified spectra⁵. Spectral library searches, introduced by Yates et al.⁶ became important with the availability of millions of confidently identified MS/MS spectra. MS library search tools include SpectraST⁷, NIST MSepSearch (<http://peptide.nist.gov/>), BiblioSpec⁸, X!Hunter⁹ and ProMEX¹⁰.

Theoretically, de novo techniques should have advantages over database search algorithms, because they can identify protein variants not contained in a database, but database search algorithms have historically outperformed de novo algorithms¹¹. However, the development of improved resolution and mass accuracy methods such as high-energy collisional dissociation (HCD) has re-ignited research on de novo methods (pNovo¹²). Hybrid similarity search- de novo methods have also been suggested^{13,14}.

A current challenge in proteomics is to rapidly identify modified peptides in complex mixtures. As powerful tandem MS becomes available to more researchers and many basic problems in identifying peptides have been resolved, the ability to identify peptides and infer protein sequences has become commonplace. The attention of the proteomics community is now focused on quantification using a variety of methods based on peak area of precursor ions, selective measurement of precursor-fragment transitions, fragmentation of all precursor ions (e.g. MS^E from Waters Corp.) or spectrum counting. All of these quantitative methods use peptide ions as measurable proxies for the protein(s) they represent and the correct assignment of peptide sequence to a spectrum is critical. Unexpected modifications can limit identification of peptides and further confound quantification as unidentified modified forms of an identified peptide will reduce the measured signal.

1
2
3 Identification of peptide modifications is of tremendous importance not only for better
4 identification of peptides but also in its own right. Post-translational modifications (PTMs) of
5 proteins play a central role in regulating their cellular function. Proteins are also modified
6 outside the cell, for example during industrial food processing and gastrointestinal digestion.
7 These modifications have important implications for human health, nutrition and allergenic
8 responses. Peptides may also be modified during sample preparation (i.e. extraction and
9 purification of proteins) or miscleaved by proteolytic enzymes prior to analysis by mass
10 spectrometry. Here we use the term PTM as a proxy for *any* peptide modification regardless of
11 its origin. Hundreds of different PTMs have been identified so far; the majority are simple ones
12 such as phosphorylation¹⁵⁻¹⁷ and it is likely that many more remain to be discovered¹⁸.
13
14
15
16
17
18
19
20
21
22
23
24
25
26

27 Specific PTM identification techniques are often developed by adjusting general peptide
28 sequence assignment tools. Database searches are widely-used, generally achieved by a
29 procedure first proposed by Yates et al.⁶ where users pre-specify which modifications are
30 expected. Specifying several modifications increases the search space exponentially and
31 increases the potential for false discoveries (PTMFinder¹⁸, PeaksPTM¹⁹). MASCOT deals with
32 this problem by a two stage process to identify unspecified modifications, called 'Error Tolerant
33 Searches' (ETS): (i) spectra are used for a database search, (ii) peptides robustly identified in the
34 first stage are used for a more relaxed search using all mass shifts that are specified in a
35 modification database. The limitations of the ETS approach are: (a) the protein must have at least
36 one high scoring peptide in the first stage and (b) the mass of the modification must be specified
37 to the search algorithm. These limitations represent a critical bottleneck in the interpretation of
38 peptide mass spectra. Granholm et al.²⁰ identified PTMs as a problem leading to incorrectly
39 assigned spectra. Other PTM finders use a sequence tag based search where de novo sequence
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

tags are computed from MS/MS spectra and used to find matches in a sequence database²¹⁻²⁴. However, this method to reduce the search space is largely limited to known protein sequences and modifications and can lead to misinterpretation if they occur inside the sequence tag. InsPect²⁵ (a database filtering method) aims to identify any PTM (i.e. is unrestricted in mass search), but needs the identification of unmodified peptides as a prerequisite. PeaksPTM¹⁹ is a newer method outperforming InsPect and MASCOT in identification of known PTMs, but is restricted to all known PTMs and needs identification of unmodified peptides as well. SeMoP²⁶ is another unrestricted PTM identifier that needs (unmodified) peptide identification first. PTM identification remains a focus of intense research²⁷.

Analysing the similarities of measured spectra has been suggested as an alternative approach for unrestricted PTM identification. Falkner et al.²⁸ developed a specific dot product similarity score for clustering spectra. However, this over-all spectra similarity score might become corrupted by the presence of noise peaks. Tsur et al.²⁹ developed a spectra alignment approach that is potentially more robust to noise and allows PTM identification within modification masses from -100 to +160 Da. Spectra alignment was later used in PTMFinder¹⁸ allowing the identification of one unrestricted PTM (± 250 Da) per peptide, and in spectral network analysis (SNA³⁰), yielding unrestricted PTM identification, but only for precursor ion charges 1 and 2.

Here we present a novel approach called Mass Spectrometry-Peak Shift Analysis (MS-PSA) for PTM identification that is not restricted to any modification masses or ion charges and is complementary to current standard methods such as MASCOT and SEQUEST. MS-PSA is not based on database search or peptide identification, but allows the identification of spectra that are related, for instance due to one or more PTMs. MS-PSA is inspired by frequent pattern mining³¹. Such bi-clustering approaches^{32,33} avoid the noise dependence of whole-spectra clustering. MS-

PSA is straightforward and fast, essentially based on frequent peak shifts that match the precursor ion mass difference. It primarily provides potential PTM masses that can be confirmed by existing standard methods. MS-PSA improves the speed and sensitivity of peptide identifications and allows unrestricted PTM identification. We present the analysis of a simple and a complex proteomic dataset to illustrate the use of MS-PSA.

MS-PSA identifies related spectra without reference to either a sequence database or a modification database. However, to aid interpretation MS-PSA permits the adoption of peptide annotations from a database search of the user's choice. The parameters used by MS-PSA are readily adjustable for analysis of data from low (e.g. ion traps) or high mass accuracy instruments at both, precursor and fragment levels (to permit the use of hybrid instruments such as the Orbitrap family). In particular, if the approximate location of a modification on a specific protein (i.e. the peptide) is suspected, MS-PSA can rapidly guide the user to informative spectra. We note that the application of the general approach used by MS-PSA is not limited to peptides and could be used for the analysis of any set of fragmentation patterns.

Methods

Mass Spectrometry-Peak Shift Analysis (MS-PSA) takes as input MS/MS spectra in MASCOT generic format. If some spectra have already been identified (e.g. by SEQUEST or MASCOT) a corresponding list containing spectrum titles, peptide annotations and modifications can be provided, but MS-PSA itself works without any annotation information. If annotation is available it assists with interpretation of results. MS-PSA outputs a ranked list of related spectra and a corresponding graph visualization.

MS-PSA workflow

The main parameters are discussed and default settings are given in brackets. Fig.S1 shows the default setting of all parameters.

1. Spectra pre-processing.

Remove spectra with very low numbers of peaks (likely noise, default: spectra with <10 peaks removed), round precursor ion masses and MS/MS peak positions according to the resolution of the MS machine used (default: precursor ion mass to 0.5 Da and fragment mass to 0.5 m/z).

Noise filtering: Keep only the $t1$ most intense peaks per fragment-mass-difference $t2$ (default: $t1=5$, $t2=100$ m/z).

2. Group spectra into groups with the same precursor mass.

3. Analyse spectra of a given group for additional group splitting.

Two different peptides might have the same precursor mass, but are still distinguishable by different peak patterns. Generally, significantly different peak patterns can be found by appropriate (bi-)clustering. Dot products have been used for clustering of MS data^{5,27,34}. Bi-clustering methods such as DASS^{32,33} group spectra with similar subsets of peaks together, such that noise peaks tend to be less distracting. MS-PSA has implemented a local optimization clustering method, using specific significance tests. It can easily be adjusted by several parameters. Clustering increases the number of spectra groups to consider. A graph output of split spectra groups helps manual revision of the clustering.

4. Remove spectra that might not belong to the group considered.

For each spectra group the *nmfp* most frequent peaks are identified. Only spectra

containing at least $mofp \leq nmfp$ such peaks are considered for further analyses and only groups containing at least $mns pg$ spectra are further analysed.

5. Identification of one ‘fingerprint’ per spectra group.

A ‘fingerprint’ is the largest closed set (cs) with $\text{frequency} \geq csf$ (default: $csf=2$). A cs is a subset of elements (in our context peak positions) that occurs more frequently in host sets (all spectra in a group) than any of its supersets^{32,33}. This represents an information rich (many peaks) but still reliable (found in at least csf spectra) set of peaks representing the group. We implemented $csf \in \{1, 1.5, 2\}$, with the following specificities: if $csf=1$ and different cs have the same maximum size, then the cs corresponding to peaks with higher intensity is chosen; if $csf=2$ and different cs have the same maximum size, then the union of the cs is taken if this is supported by at least one spectrum. If $csf=1.5$, first the most frequent peaks in the whole spectra group are calculated. All peaks with $\text{frequency} \geq peakfreq$ are considered. The fingerprint is the largest intersection of these peaks with any spectrum in the group. If $csf=1.5$ and $peakfreq=2$, the corresponding fingerprint is typically larger than for $csf=2$ and smaller than for $csf=1$; the confidence of the peaks is in between too. If different $csf=1.5$ fingerprints have the same maximum size, the one corresponding to peaks with higher intensity is chosen.

6. The core of MS-PSA: for all pairs of spectra groups potential peak shifts are identified.

If the fingerprints of two groups have at least cp peaks in common and the number of common peaks divided by the size of the smaller fingerprint is at least $qual2$, the following is done: All positive differences in fragment peak positions from the fingerprint of the lighter precursor ion to the fingerprint of the heavier one are calculated (positive difference means: a peak position of the heavier precursor ion must represent a higher

mass than a peak position of the lighter one to be considered for the difference). All differences are collected and sorted by frequency. Two spectra groups (peptides) are considered to be potentially related (by one modification, e.g. PTM) if one of the peak differences with a frequency of at least *pwss* (peaks with same shift) and a quality of at least *qual1* (*qual1*=*pwss*/size of smaller fingerprint) is deviating less than *maxdev* from the corresponding precursor ion mass differences (default: *cp*=6, *pwss*=5, *qual1*=0.1, *qual2*=0.15, *maxdev*=1 (Da, should be adjusted according to MS instrument accuracy)). The list containing the *pwss* frequent peak differences is called a peak-shift list. It can also be used for the identification of more than one potential PTM: if the sum of two frequent peak shifts is (nearly) identical to the precursor ion mass difference it could be due to two modifications. Identification of higher numbers of modifications is technically possible but not considered so far. The parameter setting the maximum number of peptide modifications to consider is *nummods* (default: 2). Note that the core of MS-PSA does not consider peak intensities, but only matches peak positions.

MS-PSA outputs two lists of related spectra groups: ‘few’ contains only the spectra containing the fingerprints of the two spectra groups (i.e. the most reliable spectra for the considered spectrum relation), ‘more’ contains all spectra of the two groups. ‘Few’ is best used when one is hunting for new modifications; this table identifies the most representative spectra assigned to the groups. The output ‘more’ is appropriate for spectrum counting or when checking spectra within peak areas assigned to peptide identifications. Sometimes peak differences deviate only by small amounts from each other, so MS-PSA also provides merged-peak-shifts: two peak shifts are merged if they deviate by ≤ 1 . Note that a manageable number of spectra groups and peak

1
2
3 numbers can always be obtained by appropriate adjustment of the parameters discussed above.
4
5 Accordingly, MS-PSA can handle MS/MS data of any size and mass accuracy.
6
7

8 The main result of MS-PSA analysis is the ‘few’ list of related spectra (Fig 1a) that can be
9 ranked by the size of the corresponding precursor ion mass difference (i.e. mass of the potential
10 modification) or by different quality parameters, such as the confidence scores *qual1/2* and the
11 corresponding sum. In addition, MS-PSA provides a graphical output of the spectral relations
12 (supplemental information, Fig. S2 and example data MS-PSA_example.zip). Note that the main
13 results table (‘few’) contains all information to visualize the spectra relations graph using
14 Cytoscape (Fig.1b); to simplify node annotation an additional annotation table is provided. To
15 enable easier interpretation of results, MS-PSA also provides spectra visualization, highlighting
16 fingerprint peaks and peaks shifted due to the potential PTM (Fig. 2e).
17
18
19
20
21
22
23
24
25
26
27
28

29 The complete Mathematica code is freely available as MS-PSA.nb (contained in the
30 supplementary file MS-PSA_example.zip). If a user of MS-PSA.nb puts the two files
31 spectra.mgf and Mascot-annotations.txt (contained in the supplementary file MS-
32 PSA_example.zip) into the folder “MS-PSA_example” and this folder is just in C, then the user
33 can run MS-PSA.nb without any modification of the code and gets exactly all results that are in
34 the folder “2014_3_10_11_57” (contained in the supplementary file MS-PSA_example.zip).
35
36
37
38
39
40
41
42

43 *Details of test set data*

44
45 We demonstrate the application of MS-PSA using two MS/MS example datasets (based on
46 collision-induced dissociation fragmentation):
47
48
49

50 (1) A single modified protein digest from Traka et al.³⁵ containing 2211 spectra, collected on a
51 LTQ (Thermo Scientific). Tandem mass spectra were extracted by Xcalibur (Thermo Scientific)
52 Charge state deconvolution and deisotoping were not performed. All MS/MS samples were
53
54
55
56
57
58
59
60

analyzed using MASCOT (Matrix Science, London, UK) to search the SwissProt/TREMBL database (selected for Homo sapiens, 108143 entries) specifying the digestion enzyme trypsin and with a fragment ion mass tolerance of 0.80 Da and a precursor ion tolerance of 1.5 Da. Iodoacetamide derivative of cysteine was specified as a fixed modification and error tolerant modifications permitted. A second Mascot search included the custom sulphoraphane (SF) modification of +217 as a variable modification. Mascot results are compiled in Scaffold (version Scaffold_4.3.0, Proteome Software Inc., Portland, OR) and presented in the supplemental information. A free Scaffold viewer is available from Proteome Software. However, note that no additional software like Mascot or a Scaffold viewer is required to run MS-PSA.

(2) A standard protein mix from Klimek et al.³⁶ (specifically “mix7 Orbitrap”); we downloaded the 10 raw files and made a single mgf file using Progenesis software (Non Linear Dynamics), provided in the supplemental information (Large_example_data.zip), in MASCOT generic format. The data were generated on a high mass accuracy instrument (Orbitrap) and therefore a precursor mass tolerance of 0.01 was allowed for the precursor ions. The standard sample contained 18 proteins and a number of known contaminants. MS/MS spectra were searched using Swiss Prot/TREMBL (17035495 entries) with the digestion enzyme trypsin and a fragment ion mass tolerance of 0.80 Da and a precursor ion tolerance of 10.0 PPM. Iodoacetamide derivative of cysteine was specified in Mascot as a fixed modification. Oxidation of methionine was specified in Mascot as a variable modification. An error tolerant search was performed with these same parameters and results are available in a scaffold file in the supplementary information (Large_example_data.zip).

Networks were visualized using Cytoscape³⁷.

Results

Analysis of a small low-resolution dataset

To test the ability of MS-PSA to assist in the identification of unspecified modifications, we re-analysed a previously published dataset³⁵. It contains a novel modification that was extremely difficult to identify for the initial publication. The sample consists of purified TGF β treated with an isothiocyanate, sulforaphane. During alkylation for tryptic digestion, an unexpected secondary chemical reaction altered the expected mass shift from +177 to +217. The Traka dataset contains 2211 MS/MS spectra, 529 of which were assigned to 14 unique peptides with an error tolerant MASCOT search (counting mis-cleavages as unique hits and including the +217 modification).

Using the default MS-PSA parameters given in Fig. S1, analysis with MS-PSA gave 83 spectral relations containing 111 unique spectra in the main results ('few') table and 699 in the 'more' table (supplemental information, Tab.S1_83_spectra_relations.xlsx and example data MS-PSA_example.zip). We propagated the MASCOT ETS peptide annotations for identified spectra on to these relations. The majority of the mass shifts represent common modifications; oxidation (+16), salt adducts (+22) or formylation (+26). Amongst the remaining mass shifts is the +217 that Traka et al.³⁵ assigned to the further modification of sulforaphane with iodoacetamide. This peptide is identified as ALDTNYCFSSTEK by the propagated MASCOT annotation. The analysis with MS-PSA found the same spectra as Traka et al., and additionally found a second group that had a salt adduct (Fig 1a). When +217 was added to the MASCOT search parameters these spectra were also identified by MASCOT.

To facilitate the interpretation of the spectral groups the main results table can be visualised as a network (e.g. using Cytoscape), where each spectral group can be represented as a node and

spectra relations (rows in Fig.1a) as edges. Edges can be labelled with mass shifts (as in Fig.1b) or other values from the main results table. Each row of the results table creates an edge between two nodes. For convenience MS-PSA provides graph pictures for the main results table and for all spectra of related groups (shown in the example dataset and Fig. S2). Additionally, MS-PSA generates a table to facilitate the labelling of nodes in Cytoscape with MASCOT annotations or the precursor mass (if no MASCOT annotation available). Thus, the 83 spectra relations could be reduced to seven networks (Fig 1b). Each node corresponds to one spectra group. Note that the same peptide might correspond to several spectra groups with slightly different parent ion masses. We grouped parent ion masses to 0.5 Da in this analysis, and therefore multiple overlaid nodes can be seen with spectra outside this mass tolerance (due to C13 isotopes or other factors). Such redundant nodes are shown as stacks in Fig.1b. The graphs in Fig.1b clearly link related spectra and the mass shifts associated with oxidation (16) and sodium adducts (22) are apparent. Miscalculation and large truncations are also obvious, such as the peptides CVPQALEPIVYYVGR and its non-tryptic cleavage product ALEPIVYYVGR, linked by 484.

If the peptide of interest is known, the graphs can quickly guide users to related spectra. We illustrate the use of the graph in Fig. 1b where we highlight the spectral relations associated with ALDTNYCFSSTEK, the peptide modified by +217. This graph clearly identifies a loss of the N-terminal alanine residue (-71), the +217 and its salt adduct (+22). Thus, this feature can assist when multiple modifications are present, by narrowing the likely sums that could account for the mass shift of more than one modification.

To assist manual verification that the two spectral groups are indeed related, MS-PSA returns the fragment masses that were used as closed sets in each group, as well as the corresponding shifted peaks. This information is returned to the user in text files, where the line number is the

1
2
3 same as the ID in the main results table. MS-PSA also provides a visualisation of all spectra
4 contained in the main results table (provided in a folder called 'spectra-pics', cf. supplementary
5 information, MS-PSA_example.zip) as well as the fingerprint peaks of the two related spectra
6 groups in one figure (in a folder called 'peak-shift-pics' in the example dataset and illustrated in
7 Fig. 2e). When combined with an annotated spectrum, this makes a powerful tool to manually
8 verify the MS-PSA assigned relationship. We illustrate this feature by comparing the annotated
9 spectra for the peptide ALDTNYCFSSTEK with and without the +217 modification (published
10 in Traka et al.³⁵) with the peak shifts identified by MS-PSA (Fig. 2). As the modification to
11 ALDTNYCFSSTEK is at the N-terminus of the peptide, all the b-ions shift, while the y-ions are
12 common. All of the peaks which MS-PSA identifies as shifted by 217 are b-ions; the y-ion series
13 remains unchanged.
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28

29 *Analysis of a large high-resolution dataset*

30
31 To test the performance of MS-PSA to identify unspecified modifications in a more complex
32 mixture, we re-analysed a standard dataset from Klimek et al.³⁶. We performed two types of
33 MASCOT search (simple specified modifications and error tolerant). These two searches enabled
34 us to ask two questions: A. Could MS-PSA identify the same modified-unmodified peptide pairs
35 as MASCOT? B. How many additional spectra could be assigned to related groups through
36 identification of additional modifications? It should be noted that Dataset 2 was acquired with
37 higher mass accuracy of the precursor ion and this was accommodated in MS-PSA by narrowing
38 the precursor mass tolerance in the parameters. The mass accuracy of the fragment spectra
39 remains the same as in the previous dataset.
40
41
42
43
44
45
46
47
48
49
50
51
52

53 Dataset 2 contains 35,014 MS/MS spectra, including many from precursor ions with charge
54 states greater than four. A simple MASCOT search (permitting only variable oxidation of
55
56
57
58
59
60

methionine residues and fixed carbamidomethyl of cysteine residues as modifications) assigned peptide identifications to 1889 spectra. Repeating the MASCOT search with error tolerant parameters assigned identifications to a further 1040 spectra. Altogether 2929 spectra were assigned to 427 peptides by MASCOT (Tab.S2_2929_Mascot-annotated_spectra.xlsx, MASCOT parameters are specified in the methods section).

To test the performance of MS-PSA we created a reference set of related peptides from these two MASCOT searches. Peptides were defined as related if (i) the same peptide was present in the MASCOT results differing by modification with a mass of greater than 15 Da, or if (ii) peptide sequences were related by mis-cleavage or non-tryptic cleavage at either end of a peptide sequence only. This definition of peptide relations excluded internal overlap of sequences or modification and miscleavage with no intermediary forms. We call this subset ‘PAIRED’ and use it to answer question A: to determine if MS-PSA could find the same mass shifts as indicated by MASCOT results. The PAIRED set contains 310 peptide relations, that can be linked into 99 graphs of related peptides in a network view of the relations (Tab.S3_310_Mascot-peptide_relations). This set covered 2102 spectra, the majority of all spectra (2929) assigned to peptide sequences by MASCOT.

Analysis of the full 35 k dataset with MS-PSA created 705 spectra relations (parameters used are given in the supplemental information), covering 705 unique spectra in the ‘few’ and 4814 spectra in the ‘more’ table. Several of these 705 spectra relations correspond to the same peptide relation, so overall MS-PSA identified 602 different peptide relations that form 153 graphs of related spectra in a network view (Tab.S4_705_MS-PSA-spectra_relations.xlsx). The overlap between the 310 MASCOT relations and the 602 MS-PSA relations is 56. The statistical significance of this overlap is beyond the working precision of standard calculation software,

1
2
3 $p < 10^{-300}$ (background of $\sim (30k)^2$ relations). Note that relaxing for instance the quality parameters
4
5 of MS-PSA leads to identification of many more potential spectra relations, finally covering all
6
7 310 PAIRED relations. As with the first test dataset, commonly identified modifications included
8
9 oxidation, sodium adducts and formylation and miscleavages.
10
11

12
13 There are 36 additional MS-PSA peptide relations to which MASCOT had also assigned
14
15 peptide identifications in both groups but these peptide pairs had not been included in the
16
17 PAIRED test set (according to the definitions given above). In all these cases MASCOT assigned
18
19 the same or very similar peptides to both groups. Examination of these showed that amino acid
20
21 substitution or variation in the assigned peptides was common, or that the sequence overlap was
22
23 internal so these did not fit our selection criteria. However, this illustrates the strength of MS-
24
25 PSA in reliably grouping spectra that should be assigned to the same peptide or modifications
26
27 thereof.
28
29
30

31
32 Combining the networks of peptide relations in the MASCOT PAIRED results and the MS-PSA
33
34 results gave a total of 181 related peptide graphs. 38 peptide graphs have at least one relationship
35
36 in both the PAIRED test set and the set identified by MS-PSA. The 56 overlapping relations are
37
38 shown in green. (Fig. S3).
39
40

41
42 Whilst the peptide relation graphs can appear to be very complicated, it should be noted that
43
44 these represent structured groups of thousands of spectra. To illustrate the utility of network
45
46 relations we consider some specific examples in Figure 3. The peptide ICGGWQMEEADDWLR
47
48 has an oxidized form, an oxidized with formylkynurenin (+31.99) form (giving a mass shift of
49
50 +80), and a tryptic mis-cleavage (ADDWLR) in the MASCOT results and PAIRED test set.
51
52 Analysis with MS-PSA identified the relation between the oxidized and mis-cleaved form (green
53
54 lines Fig. 3A) but did not assign the +80 relation (yellow line Fig. 3A). In addition MS-PSA
55
56
57
58
59
60

identified other groups of spectra for the truncated form and likely oxidation thereof (blue lines Fig. 3A). The peptide MAEEAGEENFFIFGMR also has a tryptic mis-cleavage (to AGEENFFIFGMR) that is also observed as an oxidized form. MS-PSA identified these relations and additionally other spectra for the mis-cleaved form and two other groups of spectra that are related to both MAEEAGEENFFIFGMR and AGEENFFIFGMR by mass shifts of 846, 434 and 386 and 26 (respectively see Fig 3B). These dual assignments by MS-PSA make the relationships highly credible. Moreover, the MS-PSA confidence score, represented by the thickness of blue and green lines, also helps to identify related spectra.

Discussion

According to Banderia et al.³⁰, the MacCoss group³⁸ were the first to realise the potential of spectral pairs for the identification of PTMs. The unrestricted PTM identification approaches ModifiComb³⁹ and spectra network analysis (SNA³⁰) are based on this idea. However, SNA works only for precursor ion charges 1 and 2, and ModifiComb needs the identification of unmodified peptides as a prerequisite and requires high mass accuracy data. Both methods also provide the modification site. ModifiComb and SNA work with potential b and y ion tags. ModifiComb calculates a number of common cleavage sites present in two peptides as the central parameter to quantify the peptide relatedness. SNA quantifies spectra similarity by an advanced spectra alignment algorithm. Both ModifiComb and SNA work with complete spectra, whereas MS-PSA is based on the representative fingerprints of spectra groups, making it robust to noise. ModifiComb additionally uses retention times as a second dimension, whereas Bandeira et al.³⁰ argue that “retention time analysis imposes the constraint that both spectra must come from the same sample”, so they do not consider it. We agree with Banderia et al. and do not use the retention times of the spectra in MS-PSA analysis. MS-PSA is exclusively built on the spectral pair paradigm. It has no

1
2
3 restrictions to charge states or PTM masses, nor to identification of unmodified peptides. MS-
4
5 PSA uses a completely different approach to ModifiComb and SNA, it provides a
6
7 complementary method for PTM identification. Potential PTMs that score highly in MS-PSA can
8
9 additionally be analyzed by existing technology such as MASCOT or SEQUEST, by adding the
10
11 modification mass identified by MS-PSA.
12
13

14
15 Retention time information is often used to reduce the complexity of the analysis algorithm. MS-
16
17 PSA is working efficiently without using retention time; data of any reasonable size can be
18
19 analyzed. The core of the MS-PSA algorithm (step 6 of the workflow, cf. Methods) analyses all
20
21 $n(n-1)/2$ fingerprint spectra pairs of the n spectra groups considered, the corresponding time
22
23 complexity is $O(n^2)$. The analyzed large dataset contains $n \approx 1000$ spectra groups (using the
24
25 parameters given). The complete MS-PSA analysis takes ~ 1 h on a standard PC.
26
27

28
29 MS-PSA was developed as a tool complementary to existing technology in order to facilitate the
30
31 identification of related spectra, likely representing unspecified modifications in large datasets.
32
33

34
35 Of course, our new approach has room for further improvement. Working with only the
36
37 measured spectra and no databases MS-PSA needs at least two related spectra for identification
38
39 of potential PTMs, for instance an unmodified peptide and the same with one or more
40
41 modifications, or the same peptide with one and two modifications. Providing theoretical spectra
42
43 would enable identification of potential PTMs in cases where just one modified form is present
44
45 in the spectral dataset. MS-PSA provides several quality parameters for the relation between
46
47 spectra, for instance the number of shifted peaks matching the precursor ion mass difference and
48
49 the relative number (normalized by the number of peaks of the smaller fingerprint). False
50
51 Discovery Rates (FDR) can be calculated by exploiting a decoy (i.e. appropriately shuffled)
52
53 database which should contain more spectra to enable conservative estimates; $FDR = m/n$, $m = \#$
54
55
56
57
58
59
60

identified PTMs according to the given threshold in the decoy database, $n=\#$ in measured data^{19,20}. All parameters of MS-PSA need to be set by hand in advance, a typical working default set is provided (cf. Methods, Fig.S1). However, at least some parameters could also be set automatically, depending on the data available at each step (e.g. *cp* and *pwss*, cf. Methods). The modification site of the potential PTM could be determined by the spectral products analysis method described by Bandeira et al.³⁰. For samples with reproducible liquid chromatography conditions, addition of retention time could improve results. Interestingly, a method DeltAMT was suggested for the identification of peptide modifications, that works exclusively with precursor masses and retention times (a corresponding bivariate Gaussian mixture model discriminates modification-related spectral pairs from random ones)^{40,41}. DeltAMT is fast and complementary to MS-PSA, so a combination of these approaches could provide a highly efficient method.

Finally, the MS-PSA approach may be further developed for peptide quantification. There is considerable scope for improvement of spectrum counting methods by identifying all spectra associated with a peptide. MS-PSA may also contribute to precursor area methods as an unidentified modified version of the measured peptide will subtract from the intensity of the unmodified form.

ASSOCIATED CONTENT

Supporting Information Available. Supplemental information.docx contains the parameters used for the analysis of the two test sets. Fig. S1 shows the default parameters of MS-PSA and Fig. S2 the network output from MS-PSA. Fig. S3 provides an overview of the overlapping PAIRED and MASCOT peptide graphs identified in the large test set. The two test datasets from

Traka et al.³⁵ and Klimek et al.³⁶ are provided in mascot generic format (.mgf) files. MS-PSA_example.zip contains the files MS-PSA.nb, spectra.mgf, Mascot-annotations.txt (Scaffold format .sf3) and the MS-PSA results folder 2014_3_10_11_57. MS-PSA.nb contains the complete Mathematica code. This runs without any modification needed (cf. Methods) to analyse the small test set from Traka et al.³⁵ (spectra.mgf and Mascot-annotations.txt). The large test data from Klimek et al.³⁶ is contained in Large_example_data.zip. Supplemental tables S1-S4 are provided in the four files Tab.S1_83_spectra_relations.xlsx, Tab.S2_2929_Mascot-annotated_spectra.xlsx, Tab.S3_310_Mascot-peptide_relations.xlsx, and Tab.S4_705_MS-PSA-spectra_relations.xlsx. This material is available free of charge via the Internet at <http://pubs.acs.org>.”

REFERENCES

1. Eng, J.K., A.L. McCormack and J.R. Yates, An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom*, 1994. **5**(11): p. 976-89.
2. Perkins, D.N., D.J. Pappin, D.M. Creasy and J.S. Cottrell, Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis*, 1999. **20**(18): p. 3551-67.
3. Brosch, M., S. Swamy, T. Hubbard and J. Choudhary, Comparison of Mascot and X!Tandem performance for low and high accuracy mass spectrometry and the development of an adjusted Mascot threshold. *Mol Cell Proteomics*, 2008. **7**(5): p. 962-70.
4. Tharakan, R., L. Martens, J.E. Van Eyk and D.R. Graham, OMSSAGUI: An open-source user interface component to configure and run the OMSSA search engine. *Proteomics*, 2008. **8**(12): p. 2376-8.
5. Ye, D., Y. Fu, R.X. Sun, H.P. Wang, Z.F. Yuan, H. Chi and S.M. He, Open MS/MS spectral library search to identify unanticipated post-translational modifications and increase spectral identification rate. *Bioinformatics*, 2010. **26**(12): p. i399-406.
6. Yates, J.R., 3rd, S.F. Morgan, C.L. Gatlin, P.R. Griffin and J.K. Eng, Method to compare collision-induced dissociation spectra of peptides: potential for library searching and subtractive analysis. *Anal Chem*, 1998. **70**(17): p. 3557-65.

7. Lam, H., E.W. Deutsch, J.S. Eddes, J.K. Eng, S.E. Stein and R. Aebersold, Building consensus spectral libraries for peptide identification in proteomics. *Nat Methods*, 2008. **5**(10): p. 873-5.
8. Frewen, B.E., G.E. Merrihew, C.C. Wu, W.S. Noble and M.J. MacCoss, Analysis of peptide MS/MS spectra from large-scale proteomics experiments using spectrum libraries. *Anal Chem*, 2006. **78**(16): p. 5678-84.
9. Craig, R., Cortens, J. C., Fenyo, D., Beavis, R. C., Using Annotated Peptide Mass Spectrum Libraries for Protein Identification. *J Proteome Research*, 2006. **5**: p. 1843-1849.
10. Hummel, J., M. Niemann, S. Wienkoop, W. Schulze, D. Steinhäuser, J. Selbig, D. Walther and W. Weckwerth, ProMEX: a mass spectral reference database for proteins and protein phosphorylation sites. *BMC Bioinformatics*, 2007. **8**: p. 216.
11. Ma, B. and R. Johnson, De novo sequencing and homology searching. *Mol Cell Proteomics*, 2012. **11**(2): p. O111 014902.
12. Chi, H., R.X. Sun, B. Yang, C.Q. Song, L.H. Wang, C. Liu, Y. Fu, Z.F. Yuan, H.P. Wang, S.M. He and M.Q. Dong, pNovo: de novo peptide sequencing and identification using HCD spectra. *J Proteome Res*, 2010. **9**(5): p. 2713-24.
13. Jeong, K., S. Kim, N. Bandeira and P.A. Pevzner, Gapped spectral dictionaries and their applications for database searches of tandem mass spectra. *Mol Cell Proteomics*, 2011. **10**(6): p. M110 002220.
14. Bern, M., Y. Cai and D. Goldberg, Lookup peaks: a hybrid of de novo sequencing and database search for protein identification by tandem mass spectrometry. *Anal Chem*, 2007. **79**(4): p. 1393-400.
15. Khoury, G.A., R.C. Baliban and C.A. Floudas, Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database. *Sci Rep*, 2011. **1**.
16. Creasy, D.M. and J.S. Cottrell, Unimod: Protein modifications for mass spectrometry. *Proteomics*, 2004. **4**(6): p. 1534-6.
17. Farriol-Mathis, N., J.S. Garavelli, B. Boeckmann, S. Duvaud, E. Gasteiger, A. Gateau, A.L. Veuthey and A. Bairoch, Annotation of post-translational modifications in the Swiss-Prot knowledge base. *Proteomics*, 2004. **4**(6): p. 1537-50.
18. Tanner, S., S.H. Payne, S. Dasari, Z. Shen, P.A. Wilmarth, L.L. David, W.F. Loomis, S.P. Briggs and V. Bafna, Accurate annotation of peptide modifications through unrestrictive database search. *J Proteome Res*, 2008. **7**(1): p. 170-81.
19. Han, X., L. He, L. Xin, B. Shan and B. Ma, PeaksPTM: Mass spectrometry-based identification of peptides with unspecified modifications. *J Proteome Res*, 2011. **10**(7): p. 2930-6.

20. Granholm, V. and L. Kall, Quality assessments of peptide-spectrum matches in shotgun proteomics. *Proteomics*, 2011. **11**(6): p. 1086-93.
21. Mann, M. and M. Wilm, Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal Chem*, 1994. **66**(24): p. 4390-9.
22. Tabb, D.L., A. Saraf and J.R. Yates, 3rd, GutenTag: high-throughput sequence tagging via an empirically derived fragmentation model. *Anal Chem*, 2003. **75**(23): p. 6415-21.
23. Searle, B.C., S. Dasari, M. Turner, A.P. Reddy, D. Choi, P.A. Wilmarth, A.L. McCormack, L.L. David and S.R. Nagalla, High-throughput identification of proteins and unanticipated sequence modifications using a mass-based alignment algorithm for MS/MS de novo sequencing results. *Anal Chem*, 2004. **76**(8): p. 2220-30.
24. Han, Y., B. Ma and K. Zhang, SPIDER: software for protein identification from sequence tags with de novo sequencing error. *J Bioinform Comput Biol*, 2005. **3**(3): p. 697-716.
25. Tanner, S., H. Shu, A. Frank, L.C. Wang, E. Zandi, M. Mumby, P.A. Pevzner and V. Bafna, InsPecT: identification of posttranslationally modified peptides from tandem mass spectra. *Anal Chem*, 2005. **77**(14): p. 4626-39.
26. Baumgartner, C., T. Rejtar, M. Kullolli, L.M. Akella and B.L. Karger, SeMoP: a new computational strategy for the unrestricted search for modified peptides using LC-MS/MS data. *J Proteome Res*, 2008. **7**(9): p. 4199-208.
27. Hoopmann, M.R. and R.L. Moritz, Current algorithmic solutions for peptide-based proteomics data generation and identification. *Curr Opin Biotechnol*, 2013. **24**(1): p. 31-8.
28. Falkner, J.A., J.W. Falkner, A.K. Yocum and P.C. Andrews, A spectral clustering approach to MS/MS identification of post-translational modifications. *J Proteome Res*, 2008. **7**(11): p. 4614-22.
29. Tsur, D., S. Tanner, E. Zandi, V. Bafna and P.A. Pevzner, Identification of post-translational modifications by blind search of mass spectra. *Nat Biotechnol*, 2005. **23**(12): p. 1562-7.
30. Bandeira, N., D. Tsur, A. Frank and P.A. Pevzner, Protein identification by spectral networks analysis. *Proc Natl Acad Sci U S A*, 2007. **104**(15): p. 6140-5.
31. Agrawal, R., Srikant, R., Fast Algorithms for Mining Association Rules. Proceedings of the 20th VLDB Conference Santiago Chile, 1994.
32. Hollunder, J., M. Friedel, A. Beyer, C.T. Workman and T. Wilhelm, DASS: efficient discovery and p-value calculation of substructures in unordered data. *Bioinformatics*, 2007. **23**(1): p. 77-83.
33. Hollunder, J., M. Friedel, M. Kuiper and T. Wilhelm, DASS-GUI: a user interface for identification and analysis of significant patterns in non-sequential data. *Bioinformatics*, 2010. **26**(7): p. 987-9.

34. Wang, J., J. Perez-Santiago, J.E. Katz, P. Mallick and N. Bandeira, Peptide identification from mixture tandem mass spectra. *Mol Cell Proteomics*, 2010. **9**(7): p. 1476-85.

35. Traka, M., A.V. Gasper, A. Melchini, J.R. Bacon, P.W. Needs, V. Frost, A. Chantry, A.M. Jones, C.A. Ortori, D.A. Barrett, R.Y. Ball, R.D. Mills and R.F. Mithen, Broccoli consumption interacts with GSTM1 to perturb oncogenic signalling pathways in the prostate. *PLoS One*, 2008. **3**(7): p. e2568.

36. Klimek, J., J.S. Eddes, L. Hohmann, J. Jackson, A. Peterson, S. Letarte, P.R. Gafken, J.E. Katz, P. Mallick, H. Lee, A. Schmidt, R. Ossola, J.K. Eng, R. Aebersold and D.B. Martin, The standard protein mix database: a diverse data set to assist in the production of improved Peptide and protein identification software tools. *J Proteome Res*, 2008. **7**(1): p. 96-103.

37. Shannon, P., A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, 2003. **13**(11): p. 2498-504.

38. MacCoss, M.J., W.H. McDonald, A. Saraf, R. Sadygov, J.M. Clark, J.J. Tasto, K.L. Gould, D. Wolters, M. Washburn, A. Weiss, J.I. Clark and J.R. Yates, 3rd, Shotgun identification of protein modifications from protein complexes and lens tissue. *Proc Natl Acad Sci U S A*, 2002. **99**(12): p. 7900-5.

39. Savitski, M.M., M.L. Nielsen and R.A. Zubarev, ModifiComb, a new proteomic tool for mapping substoichiometric post-translational modifications, finding novel types of modifications, and fingerprinting complex protein mixtures. *Mol Cell Proteomics*, 2006. **5**(5): p. 935-48.

40. Fu, Y., W. Jia, Z. Lu, H. Wang, Z. Yuan, H. Chi, Y. Li, L. Xiu, W. Wang, C. Liu, L. Wang, R. Sun, W. Gao, X. Qian, and S.-M. He, Efficient discovery of abundant post-translational modifications and spectral pairs using peptides mass and retention time differences. *BMC Bioinformatics*, 2009. **10**(Suppl I): S50.

41. Fu, Y., L.-Y. Xiu, W. Jia, D. Ye, R.-X. Sun, X.-H. Qian, and S.-M. He, DeltAMT: A statistical algorithm for fast detection of protein modifications from LC-MS/MS data. *Mol Cell Proteomics*, 2011. **10**.5: 10.1074/mcp.M110.000455.

Author Contributions

TW and AMEJ devised the method. AMEJ provided the MS data. TW implemented the method. TW and AMEJ analysed the data and wrote the paper. Both authors have given approval to the final version of the manuscript.

Funding Sources

TW gratefully acknowledges the support of the Biotechnology and Biological Sciences Research Council (BBSRC). This research was funded by the BBSRC Institute Strategic Programme Grant BB/J004529/1. AMEJ gratefully acknowledges support from the Gatsby Foundation and the School of Life Sciences, University of Warwick.

ACKNOWLEDGMENT

ABBREVIATIONS

MS-PSA, Mass Spectrometry-Peak Shift Analysis

Figure Legends

Figure 1 A. Example of the main result table generated by MS-PSA. Column B contains the most representative spectra of spectra group 1 (typically two spectra if $csf=2$). A spectrum is represented by its name, containing the precursor mass (first number) and the spectrum number in the mgf file (second number). Column C contains the names of the spectra representing group 2. Groups of columns A and B are related by the mass difference in column D, with the number of such peaks in column E. Columns F and H contain parameters measuring the quality of the suggested spectrum relation, Column I contains the corresponding sum. For the sake of clarity the following columns are omitted in the figure: Col. M contains the shifted peaks (peak shift list (cf. Methods) including mass and frequency), col. N the corresponding maximum frequency. Col. O indicates if the most frequent shift matches the precursor mass difference (difference between cols. E and $N=0$). Columns J-L provide information for merged peaks. The final columns provide the titles of MS2 spectra for groups 1 and group 2.

Figure 1B. The MS-PSA network graph for the small low-resolution example, visualised using Cytoscape. Nodes represent groups from the first two columns of the main results table and are labelled with the peptide sequences propagated from Mascot, or with the precursor mass if not identified by Mascot. Edges joining related spectra groups are labelled with the corresponding mass shift. The width of the edge indicates confidence, using column I of the main results table (A). Overlaid nodes indicate multiple groups with the same mass difference, a feature created by the mass tolerance specified in the specific parameters used.

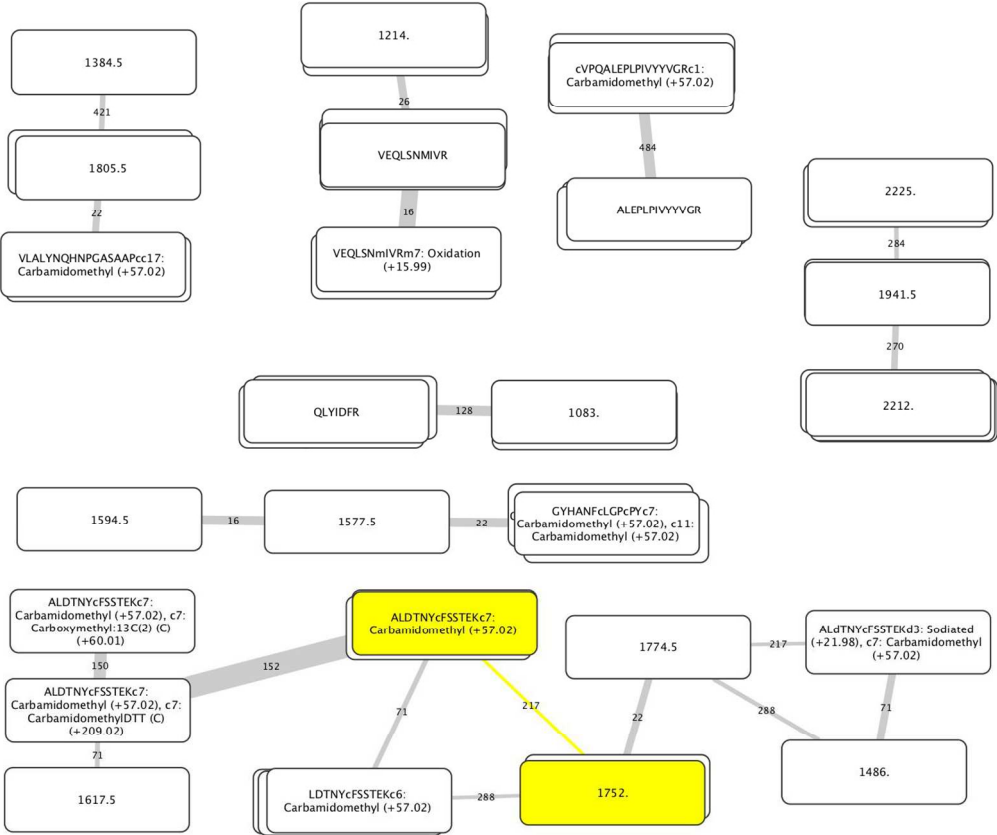
Figure 2. A-D The previously published fragmentation and the corresponding annotated spectra of the peptide ALDTNYCFSSTEK, with and without the +217 modification. The peaks marked with a triangle are the ones that shift with the modification. The shifted peaks identified by MS-PSA shown as mirrored spectra, (E) and the fingerprint sets with shifted peaks marked in red (F)

Figure 3. Examples from the large high-resolution dataset. Nodes represent peptides/spectra (labelled mass_peptide sequence, or when no Mascot annotation is available mass_MS-PSA line_MS-PSA group). Edges represent relations between peptides/spectra and are labelled with the mass shift identified. Green edges indicate that both MASCOT and MS-PSA identified the same relation, yellow is Mascot alone and blue is MS-PSA alone. The thickness of the blue and green lines represents the MS-PSA confidence score.

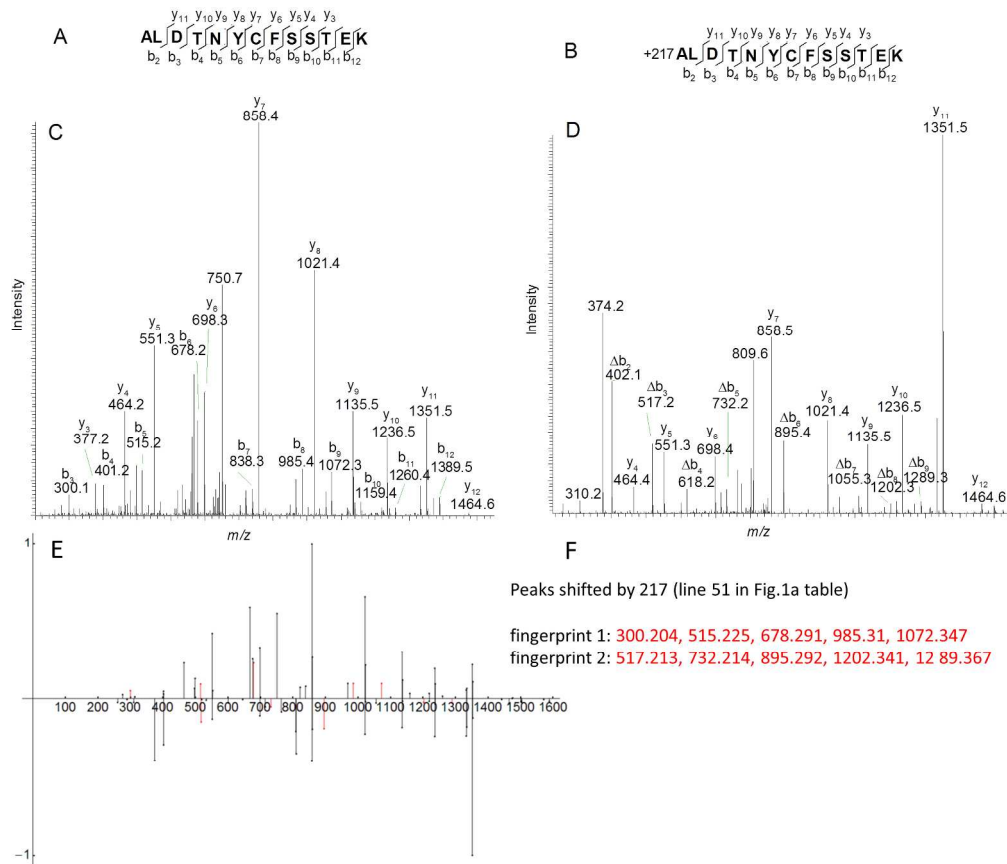
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

	A	B	C	D	E	F	G	H	I
1		MS-PSA ID specs1	specs2	peakshift matching mass differ	# such peaks	#sp/length small cs	# common peaks	#cp/length small cs	quality sum
49	48	{ "1535.--148-rm20_SFhi1.1	{ "1687.5--258-	152	14	0.341	9	0.22	0.561
50	49	{ "1536.5--211-rm20_SFhi1	{ "1687.5--258-	152	18	0.581	6	0.194	0.774
51	50	{ "1535.5--155-rm20_SFhi1	{ "1752.--768",	217	5	0.132	16	0.421	0.553
52	51	{ "1752.5--784", "1752.5--8	{ "1535.--148-r	217	5	0.122	13	0.317	0.439
53	52	{ "1557.5--165-rm20_SFhi1	{ "1774.5--786	217	5	0.132	10	0.263	0.395

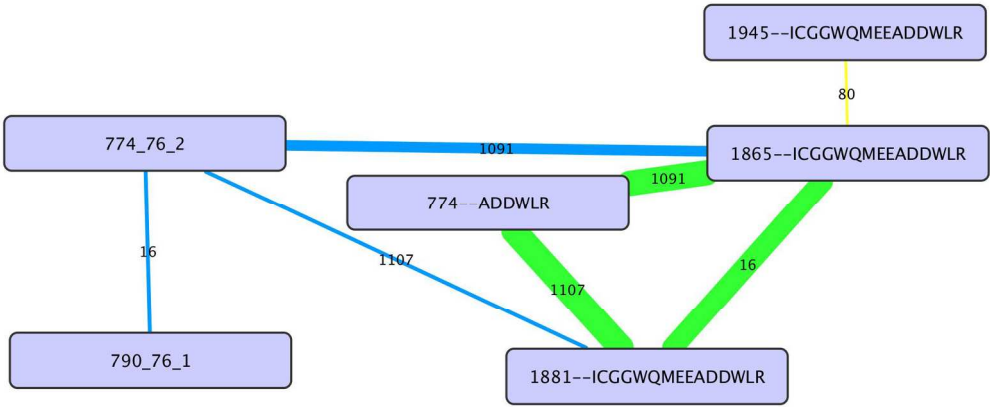
224x73mm (96 x 96 DPI)



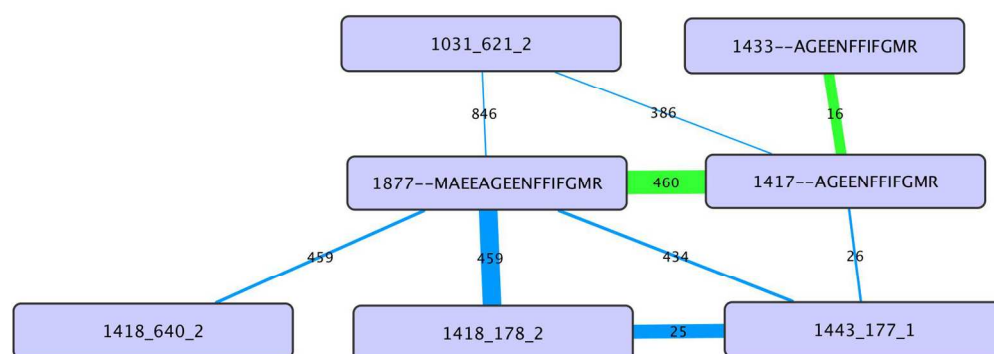
511x428mm (72 x 72 DPI)



222x190mm (300 x 300 DPI)

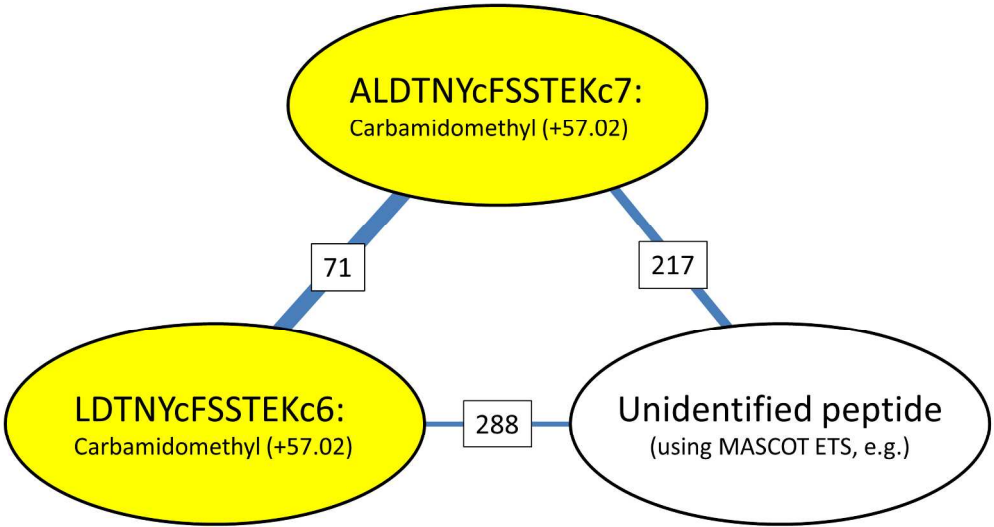


711x297mm (72 x 72 DPI)



626x236mm (72 x 72 DPI)

Peptide relations identified by MS-PSA



219x138mm (300 x 300 DPI)